IADIS International Journal on Computer Science and Information Systems
Vol. 1, No. 2, pp. 15-31
ISSN: 1646-3692

TOWARD EFFICIENT DETECTION OF CHILD PORNOGRAPHY IN THE NETWORK INFRASTRUCTURE

Asaf Shupo and Miguel Vargas Martin *University of Ontario Institute of Technology, 2000 Simcoe St. N., Oshawa, Canada, L1H7K4*

Luis Rueda *Universidad de Concepción, Víctor Lamas 1290, Concepción, Chile*

Anasuya Bulkan, Yongming Chen and Patrick C.K. Hung *University of Ontario, Institute of Technology, 2000 Simcoe St. N., Oshawa, Canada, L1H7K4*

ABSTRACT

Child pornography is an increasingly visible problem in society today. Methods currently employed to combat it may be considered primitive and inefficient, and legal and technical issues can exacerbate the problem significantly. We propose a network-based detection system that uses a stochastic weak estimator coupled with a linear classifier, which is appropriate in this context due to the non-stationarity of the input data. Our experiments show that the system is capable of distinguishing child pornography images from non-child pornography images even when the obscene image is reduced to only 20% of its representation. This method for identifying offensive material is potentially attractive to law enforcement and can be accomplished with acceptable overhead. We believe our approach, with minor adaptations, is of independent interest for use in a number of network applications which benefit from packet classification beyond detecting child pornography. These include security applications such as detecting malicious packets, and network anomalies consisting of dangerous traffic fluctuations, abusive use of certain services, and distributed denial-of-service attacks.

KEYWORDS

Computer forensics, packet classification, P2P networks.

1. INTRODUCTION

Child pornography is increasingly prevalent in today's society and steps must be taken to combat this insidious phenomenon. Even though a fine line may be crossed between freedom of expression and moral and social obligations, the fact remains that a firm stance must be

taken to avoid the negative impacts that defenseless children face due to exploitation. An example of the liberal viewpoint surrounding this issue when it comes to being lenient with child pornography and its offenders is evident in recent headlines regarding the Dutch,. It was reported that a group of pedophiles is being allowed to contest the general elections in the Netherlands in upcoming November, 2006 and they have plans to legalize the possession of child pornography and even allow it to be aired on local television [31].

This may be an immoral move that is unimaginable to the conservatively inclined, but it can serve as a catalyst to raise more awareness of the problem and its implications. At the same time, raising awareness about child pornography is not the only solution, since laws must be suitably designed to deal with offenders. Accordingly, the Canadian Constitution has been adjusted via the Bill C-2 Act which was introduced in 2004 in order to amend the Criminal Code. This was primarily done to create a broader definition of child pornography to include audio and written materials; punishment for advertising the materials; harsher penalties with mandatory minimum sentences; and a new sexual exploitation offence that protects victims aged 14 to 18 by focusing on the offence committed rather than the issue of consent.

Other improvements were the Evidence Act, which stipulates that young children under the age of 14 may testify and give evidence under oath, as well as the introduction of voyeurism charges that can be enforced when there is a violation of reasonable privacy expectations [31, 6]. There have been challenges to this new legislation and Regina v. Sharpe is one successful case where the legislation was sufficient to overcome the defence of freedom and expression [16].

While the legal framework has been responsive to the arising need to combat this problem, social programs and efforts have been increasing as well. The TPS (Toronto Police Service) contacted Microsoft Canada in 2003, which responded by delivering a software package called CETS (Child Exploitation Tracking System) worth approximately US$2.5 million [15]. However, the CETS initiative is more geared to a transparent way of dealing with large amounts of information across police services to better track suspicious activities. It is a positive approach, but not a purely technical solution to combat the problem.

Our contribution is a network-based child pornography detection system that uses well-known estimators and classifiers with appropriate adaptations to extract suspicious traffic. The accuracy and computational overhead of the system takes into account packet fragmentation and processing performance constraints inherent in routers of the network infrastructure. We test our system using child pornography images provided by the Toronto Police1 and random non-child pornography images downloaded from the Internet. We have shown that our system is capable of distinguishing between these two kinds of images with acceptable error rates even when the image is reduced to only 20% of its original representation.

2. RELATED WORK

In this section we outline relevant work related to host- and network-based approaches, including the problem of detecting censured peer-to-peer traffic. For an extensive analysis of

[1] The handling of the images adhered to the rules imposed by the Ethics Research Board of the University of Ontario Institute of Technology.

social and legal aspects refer to the work of Schell et al. [25]. The conceptual idea of this paper was presented in [8].

2.1 Host-based approaches

One area related to child pornography detection is *content-based image and video retrieval* (CBIR). This area has become more important considering the increased use and sharing of digital visual (image and video) files. The associated complexity of the problem is to be able to retrieve visual files based on their semantics. The semantics of a file are determined according to a set of characteristics (e.g., colour contrast, shapes, etc.) learned a-priori from similar files. Image and video retrieval techniques are already being proficiently exploited (e.g., [5]) and have been extensively studied (e.g., [13]).

The problem of detecting hidden messages within images has been extensively studied in the area of computer forensics. Lyu et al. [17] present a steganalysis system which analyzes images based on image representation (e.g., wavelet decomposition, local angular harmonic decomposition), magnitude (e.g., space, orientation, scale, and colour), and phase statistics. Their system then classifies images using support vector machines. The accuracy of these algorithms depends mainly on the size of the hidden message; however they were able to detect messages of reasonable size embedded in compressed JPEG.

Parental controls are used to protect children from accessing unwanted content via the Internet, whether by e-mail, instant messaging, through illegal websites, or peer-to-peer transfers. In fact, according to a study conducted on 1,100 families by the Pew Internet and American Life Project, it was revealed that 54 percent employ this mode of protection [7]. While this increased support is attributed to being the only feasible alternative available, the fact is that there are limitations that surround these Internet filters.

Numerous software packages are available on the market, and it is becoming increasingly problematic to determine which filter is robust enough to detect the highest incidence of child pornography on the Internet. CyberPatrol, Net Nanny, SurfWatch and CYBERsitter are some of the popular ones available, and usually come with a pre-determined blacklist that restricts access to content that is known to be of questionable nature. These lists can be customized to suit the child as well as the preferences of the parents, but need to be continually updated. Another approach is to allow the ISP to provide this service using their already-customized list. This requires no installation on the desktop, which may be the preferred option for some parents who are not technically inclined [24].

The other method is to list appropriate content on a white-list, allowing restricted access to that information while blocking access to all other content. Site labels with ratings are assigned to website content, and these can be used to determine access rights. Automated scanning of text, whether on a website or as part of a search query, is another technology employed to determine illicit content. A server log file can be used to store all activity records that were collected from browsing activities online. Another method is to implement passwords, encryption techniques and other authorization controls such as credit card numbers to access certain services or sites [10].

These methods do not provide complete protection, since controls can easily be circumvented by children who are sometimes more knowledgeable than their parents. Alternatively, simple search queries in Google on how to bypass parental filters can yield instructions to successfully do so. In addition, by simply misspelling words in a Google query,

suggestions and tips are shown and access to these forbidden sites is allowed. Finally, proxies which are freely available online can be used to circumvent sites or content that is blocked or placed in a blacklisted mode. The content from the blocked website is directed through the proxy to be delivered to the client that requested it [27].

Parental controls have in the past often been faulty, but are continuously improving. They are becoming increasingly visible features within operating systems such as Apple's latest version of Mac OS X, which ships with a flexible system for parental controls [4]. A popular approach to using parental controls is to begin by white-listing sites and contacts that children may access, and as the child matures move to blacklisting content instead. Another commonly suggested path is to educate children on making responsible choices on the Internet in order to protect themselves [1]. This approach, however, is often linked to social awareness programs that aim to educate society as a whole to eradicate the problem. While there is no substitute for social awareness programs, currently employed host-based solutions suffer limitations as well. Not all hosts are accessible or open to scanning, nor do all individuals know how to enable parental controls or other safeguards. A network infrastructure approach could overcome many of these issues.

2.2 Network-based approaches

The Google subpoena requiring the search giant to release its search records, allowing authorities to track access to child pornography, epitomizes this twofold issue clearly. Privacy advocates have been outraged, but at the same time recognize that the problem of swapping these images is clearly illegal and is an increasing problem.

There have been a number of important efforts to classify packets at the network level. There are intrusion detection systems, which perform packet reassembly (see e.g., [23]); however, the imposed overhead impedes the viability of these techniques in actual use within core routers. Upon the impracticability of packet reassembly, it has been shown that statistical analysis can effectively classify packets. For example, Shanmugasundaram *et al.* [26] propose a classification scheme capable of identifying the source application responsible for generating a packet. Wang *et al*. [30] offer a network intrusion detection system based on *n*-gram (an *n-gram* is a consecutive sequence of *n* characters or symbols in a text document) analysis, whereas Matrawy *et al.* [18] proposition a statistical method based on (*p, n*)-grams (a (*p, n*)-gram is an *n*-gram at position *p*). However, Zuev *et al.* [33] propose a packet classification scheme based on statistical characteristics of traffic.[2] This scheme is based on the Naïve Bayes classification technique, and is capable of classifying packets into a number of application categories, including P2P, e-mail, and multimedia files. Moore *et al*. [19] present a packet classification scheme, which incorporates a number of classification methods into a single system that combines them according to certain rules. Although these two classification strategies use higher order Markovian models, they are prone to ignore data distributions that quickly change in time, e.g. data showing some degree of non-stationarity. Furthermore, textual files can be flagged as censured material using e-mail surveillance techniques. For example, Keila *et al.* [12] propose a system, which automatically analyzes e-mail messages to determine whether they contain deceptive or unusual material. This system is based on frequency analysis of certain pronouns, verbs, emotions, and particular words.

[2] The classification is also performed on a per-byte and per-flow basis.

2.3 Censured peer-to-peer traffic

With the arising use of file sharing through peer-to-peer (P2P) forums such as Kazaa [11] and LimeWire [14], the issue of censoring the content that is being transferred infringes on legal rulings and ethical viewpoints; particularly the First Amendment in the U.S., which dictates the right to freedom of speech. Privacy rights are inevitably trampled since it may be difficult to distinguish between effectively censoring illegal material and preserving privacy rights and privileges. In particular, a number of difficulties are presented by Allen [2]; some of the most important being to ensure that digital evidence remains unaltered, searching through thousand of files, analysis of password-protected or encrypted content, infringement of privacy rights, and a number of legal issues surrounding computer forensics.

The P2P field enjoys much attention within research circles, and it is conceivable that it may become a primary distribution point for child pornography in the future. Features such as censorship-resistant publishing make it possible to anonymously and systematically distribute sets of files on networks – an environment that can be abused for malicious purposes [28]. The problem presented by the P2P field has different possible approaches, since the various P2P protocols in existence have different methods to upload or download material. As such, targeting any single P2P network requires detailed knowledge of the protocol in use. Network infrastructure techniques can be developed to analyze traffic without knowledge of overlying protocols or the specifics of networks such as the ones created by P2P applications. Any solution targeted at a specific protocol, application, or network will have limited applicability, and likely decreasing utility as offenders move to new ways of distributing child pornography. It is important to consider the case of anonymous and encrypted file P2P applications such as Winny [32]. These applications introduce additional complexity for detection and identification. Encrypted payloads are not susceptible to CBIR or e-mail surveillance techniques. Ohzahata *et al.* [20] present identification mechanisms based on TCP handshake behaviour. Their system resides inside an autonomous system, and is capable of detecting encrypted P2P file transfers. Attribution of anonymous P2P nodes at the network infrastructure level is challenging, and is a problem yet to be addressed. Identifying encrypted traffic is out of the scope of this paper, and is an area that we plan to explore in the future.

3. DETECTING CHILD PORNOGRAPHY IN THE NETWORK INFRASTRUCTURE

Currently, law enforcement members detect child pornography by manually searching the Internet or visiting chat rooms in an attempt to apprehend suspects. This is time-consuming, requires considerable human processing power, and there is no certainty that the offenders will be caught. By employing analysis at the network level, a wider volume of traffic can be scanned and can be more precisely traced to the originating host.

More importantly, detecting malicious packets that may contain child pornography is not a straightforward task due to fragmentation that occurs at the network level, specifically at level two of the OSI model, otherwise known as the data-link layer. An MTU (Maximum Transmission Unit) of 1500 bytes is imposed at this level, increasing the likelihood that child pornography files are fragmented into a number of small packets and making analysis difficult to perform. In addition, packet assembly is not available at this level to facilitate the deduction

of whether the packet is indeed an image of child pornography or not (e.g., by using a typical host-based approach).

Due to this shortcoming at the router level, we are faced with the arduous task of determining whether any packet that passes the router is obscene (child pornography) or non-obscene (non-child pornography) based on the statistical features it possesses. Therefore, there can be false negatives (i.e., obscene packets incorrectly deemed as non-obscene) or false positives (i.e., non-obscene packets misclassified as obscene) since the analysis cannot be performed on the entire packet.

Furthermore, additional overhead will be placed on routers to classify packets and this may especially be reflected on routers which handle large volumes of traffic. As such, there would be unavoidable delays in packet forwarding. However, taking into account these notable shortcomings, it is still a highly plausible scenario that child pornography can be effectively classified at the router level by employing appropriate adjustments and filtering.

3.1 Our approach

We propose to use an estimator for feature extraction coupled with a linear classifier. The estimators we consider are the stochastic learning weak estimator (SLWE) and the maximum likelihood estimator (MLE). The SLWE is considered to be more accurate in dealing with non-stationary data (moving images or clips interspersed with text), which is important since pornography is not limited to textual conversations in an email or chat room. JPEG images are considered to be part of the visual non-stationary data since they are not limited to simply moving frames. Our transmission channel is unencrypted since encryption will increase the entropy of the data, which in turn will increase the difficulty of learning in the training phase. The classification process entails the training and testing phases.

In the training phase, we have used two vectors to classify the obscene and non-obscene packets. By extracting statistical properties from the labeled packets (i.e., those that have been sorted into the obscene and non-obscene vectors), we were able to conduct the experiments using this as a basis of comparison. The features that were extracted from the training phase with any needed adjustments were input into the classifier that we used in the validation phase of the SLWE. The experiments were repeated for the MLE (the estimators are discussed in Sec. 4).

The training phase, as pointed out above, aims first to extract the statistical features of the packets corresponding to all images in the training dataset, producing one vector for each class. The following algorithm produces these two vectors when it is run for each dataset.

Algorithm Frequencies

1. Initialize an array B of counters to zero
2. For each image I of the training dataset of class j:
    2.1 For each 8-bit byte b_j of I:
        2.1.1 Increment $B[b_j]$ by 1
3. Initialize an array V_j of probabilities to zero
4. For k = 0 to 255
    4.1 Set $V_j[k] = B[b_j]$ / total number of 8-bit bytes of the set of images.

The algorithm is used separately for the obscene and non-obscene training datasets. The output of the algorithm is a feature vector, an array V_o or V_n, one for the obscene and non-obscene dataset, respectively.

Once we have extracted the statistical characteristics into the feature vectors V_o and V_n, the next step is to use an estimator to extract the features of the image to be classified, namely a vector V'. We have tested two algorithms for this purpose, the MLE and the SLWE.

The classification rule consists of assigning an unlabeled package to the class, obscene or non-obscene, that minimizes the distance between V' and the trained arrays V_o or V_n. Four metrics have been used for this purpose, which are later explained in detail (see Sec. 5), namely the Euclidean distance, the weighted Euclidean distance, a variance approach, and the counter distance. The methods calculate the distance from the actual packet to the two labeled vectors, after which a classification is made as to whether a packet is obscene or not.

To take into consideration that child pornography images may not be completely contained in a single packet due to fragmentation, we have conducted experiments in which the images were gradually reduced from 100% to 20% of their original size. Subsequently, false negatives and false positives were recorded based on the classification results. The expectation was that the higher the percentage of the entire image that was available, the fewer misclassifications should be encountered (i.e. false positives and false negatives). To interpret our results, we arbitrarily consider that the best performance is when there is an almost equal allocation (fifty/fifty) in both domains so that there is an overall low error rate when it comes to the classification phase. However, it is important to note that this is part of an implementation goal since it depends on the nature of the system to determine if it is better to assign a higher false positive or false negative threshold. This problem can be modeled in terms of minimizing the decision risk [9], which is more general than that of minimizing the classification error.

4. ESTIMATORS

Before using a classification metric (see Sec. 5), we need to extract statistical characteristics of datasets. Next, we describe each of the estimators used to extract such features. For this purpose, we obtain the frequency for each of the symbols, from 0 to 255, for a given image that has to be classified.

4.1 The maximum likelihood estimator

The maximum likelihood estimator (MLE) is a traditional technique that aims to maximize the likelihood that a given sample generates using a specific probabilistic model, either parametric or non-parametric. We assume that we are dealing with a multinomial random variable with 256 possible realizations (one symbol for each 8-bit ASCII value). It has been shown that the likelihood is maximized when the estimate for each symbol is given by the frequency counters divided by the total number of bytes in the image [9, 29]. The algorithm can be described as follows:

MLE Algorithm

1. For each image H captured by the router:
    1.1 Initialize an array C of counters to zero
    1.2 For each 8-bit byte b_j of H:
        1.2.1 Increment $C[b_j]$ by 1
2. Initialize an array V' of probabilities to zero
3. For k = 0 to 255
    3.1 Set $V'[k] = C[b_j]$ / total number of 8-bit bytes of this image.

The algorithm produces an array *V'* that contains the estimates for each 8-bit byte in the testing image *H*. That vector *V'* is then input to the classification rule, which decides on the class based on a distance function and the trained feature vectors.

4.2 The SLWE algorithm

Estimators like the one described by the MLE algorithm (Sec. 4.1) suffer from a lack of ability to capture quick changes in the distribution of the source data, e.g. dealing with non-stationary data, that is, data from different types of scenarios. Oommen *et al.* [21] proposed a *stochastic learning weak estimator* (SLWE). The SLWE combined with a linear classifier has been successfully used to deal with problems that involve non-stationary data and has been effectively used to classify television news into business and sports news [18].

In our context, each image to be classified is read from the testing dataset, and is used to feed the classification rule by means of extracting statistical features into a feature vector. The source alphabet contains n symbols (n = 256), which represent the possible realizations of a multinomial random variable, and whose estimates are to be updated by using the SLWE rules. While this rule requires a "learning" parameter, λ, it has been found [22] that a good value for multinomial scenarios should be close to 1, e.g. λ= 0.999. The algorithm is described as follows:

SLWE Algorithm
1. For each image *H* captured by a router:
   1.1 Initialize each entry of the feature vector *V'* to 1/256
1.2 For each 8-bit byte b_i of *H*:
   1.2.1 For k = 0 to 255
      1.2.1.1 If $i \neq b_i$ then
            $V'[k] = \lambda * V'[k]$
         Else
      $V'[b_i] = V'[b_i] + (1-\lambda) \Sigma V'[k]$ (for $k \neq i$)

The classification rule is validated using labeled images, and adjustments are made if necessary. Note that in the actual classification process the label of each image is not known. To classify the complete image we use four different distance functions, which are described in the following section.

5. CLASSIFICATION DISTANCES

The choice of a distance function (also referred to as "metric") is not a trivial task. Often, different components of the feature vectors may have different weights in classification of an arbitrary image. Some entries of the feature vector may be more important than other entries, or some entries may have more noise than other entries. Therefore, the choice of a metric plays an important role in the performance of the algorithm.

It is not easy to see why a distance function performs better than another. The distance that yields the best results in our experiments will be chosen. In this paper the distances between the feature vector of an arbitrary image and the feature vectors of child pornography images and non-child pornography images is based on four well-known distance metrics.

5.1 Euclidean distance

In this metric, it is assumed that all entries in the feature vector have equal weight. The Euclidean distance between two feature vectors *V* and *V'* is defined by the following equation:

$$d(V,V') = \sqrt{\sum_{i=0}^{255}(V[i]-V'[i])^2}$$

5.2 Weighted Euclidean distance

This distance is also known as the Mahalanobis distance when the covariance matrix is considered as a diagonal matrix. We suppose that different entries in the feature vector have different importance in classifying images. We also consider an entry in the feature vector of less importance than another entry if its variance is greater than the variance of another entry. We define the weighted factor *w* as $w = 1/\sigma^2$, and the distance by:

$$d(V,V') = \sqrt{\sum_{i=0}^{255}\frac{(V[i]-V'[i])^2}{\sigma^2}}$$

In the Appendix, we discuss some issues related to the weighted factor *w*.

5.3 Variational distance

This distance is usually named as *variational* distance when *V* and *V'* represent probability distributions, and *L1* distance or *city block* distance when *V* and *V'* are considered as vectors of *n*-dimensional space. In our particular case, this distance is calculated as follows:

$$d(V,V') = \sqrt{\sum_{i=0}^{255}|V[i]-V'[i]|}$$

5.4 Counter distance

In this metric we define the distance of a test vector *T* from two fixed vectors *V* and *V'* by:

d(*V, T*) = number of elements for which:

$$|V[i]-T[i]| < |V'[i]-T[i]|$$

d(*V', T*) = number of elements for which:

$$|V[i]-T[i]| \geq |V'[i]-T[i]|$$

6. EXPERIMENTS

6.1 Experimental datasets

We gathered our dataset by using sanitized child pornography images obtained from the Toronto Police, while the non-child pornography dataset was assembled using random images obtained from the Internet. We remark that the research team adhered to the Ethics Research Board of our Institution for the handling of the datasets.

Our datasets consist of two sets of files: (1) The first one is a set of JPEG files of "sanitized" child pornography images provided by the Toronto Police. These sanitized images have been anonymized by distorting the faces of all the individuals that appear in the image in such a way that their identification is no longer possible. The average size of these images is 60 Kbytes. We call this dataset "obscene." The handling of this dataset observed strict rules imposed by the Research Ethics Board at our Institution. (2) The second dataset consists of non-child pornography images downloaded from arbitrary sources through the Web. These images correspond to JPEG files of legal content such as sports, etc; all of which are considered to be "non-obscene." The average size of these images is 45 Kbytes.

6.2 Methodology of experiments

To measure the accuracy of the proposed approach, we have conducted a set of experiments using the two datasets outlined in Sec. 6.1.

First, we trained the system using the "obscene" dataset. Secondly, we tested the performance of both estimators outlined in Sec. 4.1 and 4.2 using obscene and non-obscene files. Considering the fact that a single image is usually transported via a number of packets (i.e., due to fragmentation), meaning a single packet may contain only a certain portion of an image, we processed different percentages of the images starting from 100% and gradually reducing the factor down to 20%.[3]

Each image (or image portion) processed is decompressed. As for picture encoding, we are taking 8 bits as the data unit (see Sec. 4). JPEG files are compressed images where each "byte segment" contains the information needed to reconstruct the original image. For example, a byte or group of bytes may represent the encoding of some coefficient that results from the discrete cosine transform, or other transformation. In some cases, however, the encodings result in variable length codes (i.e., of length not necessarily multiple of 8) which are spread out in a number of bytes.

All experiments were performed on an IBM laptop, with a Pentium M 1.86GHz processor, 1GB RAM, running Windows XP. We are aware of the limitations and high performance demands in routers of the network infrastructure. However we believe that such experiments can still be helpful in a preliminary assessment of new and more effective methods to combat child pornography from the network level perspective.

[3] The image reduction was achieved by reducing the height coefficient, y, from 1.0 to 0.2. The resulting portion, considered for the experiments, corresponds to the upper part of the image.

6.3 Results

In general, the less of the percentage of the original image that is available, the higher is the rate of false positives. This is demonstrated in most of the graphs of Figures 1 through 4. For the SLWE, the Weighted Euclidean metric produced the best results for both the false positive and false negative rates. The lowest percentage of false positives was for this metric in comparison to the other three metrics, and this was dependent on the percentage of the image. For the MLE algorithm, the same results were recorded, but for the false negative rate the Euclidean metric yields better results. In the Appendix, we justify the parameters used in the Weighted Euclidean metric and explain why this metric produces more accurate results in most cases.

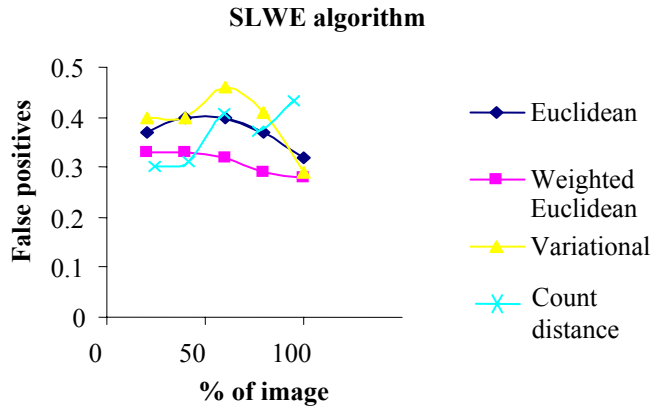


Figure 1. False positives triggered by the SLWE algorithm

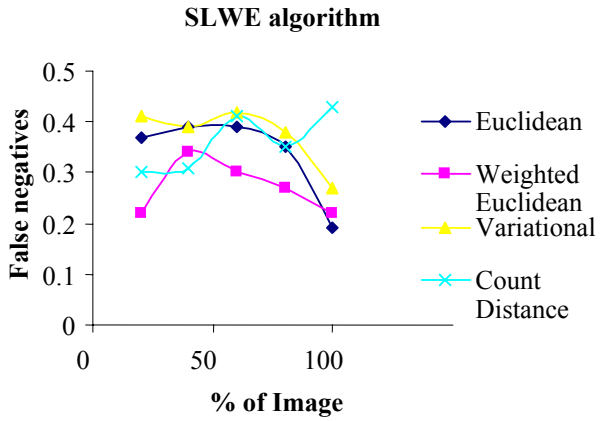


Figure 2. False negatives triggered by the SLWE algorithm

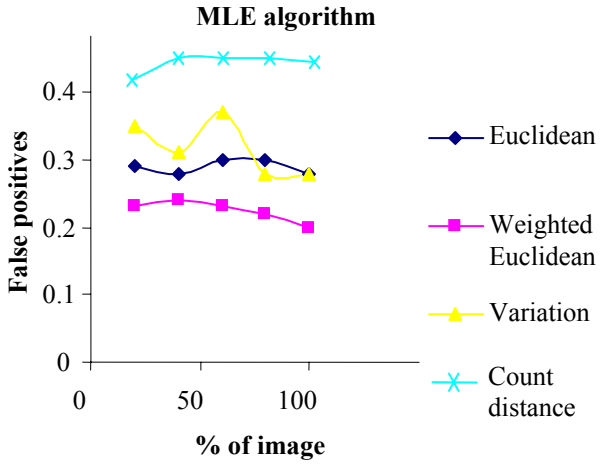


Figure 3. False positives triggered by the MLE algorithm

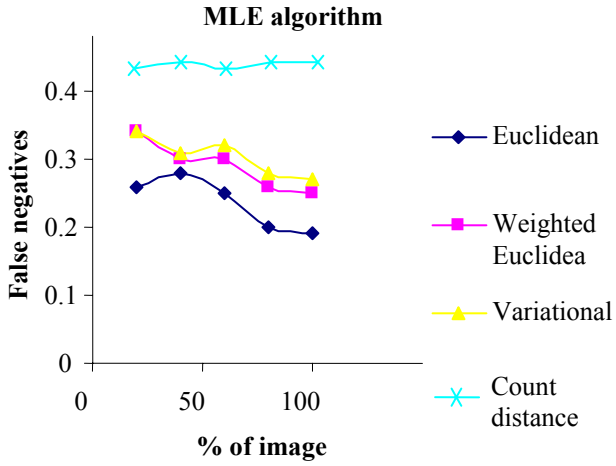


Figure 4. False negatives triggered by the MLE algorithm

Furthermore, we found that it is possible to reduce the error rate for both false negatives and false positives by considering only the third RGB component for each pixel (i.e., the blue color rate) and using the MLE estimator coupled with the Euclidean Weighted distance. Fig. 5 shows that the probability of false positives can be reduced to as much as 6.5% processing only 20% of the images. This behaviour may be due to the fact that taking all the data, that is, the full RGB, the classifier overfits the data, which is a usual problem in pattern recognition, and has to do with the classifier only. In this particular case, we found this scenario manually, by trial and error.

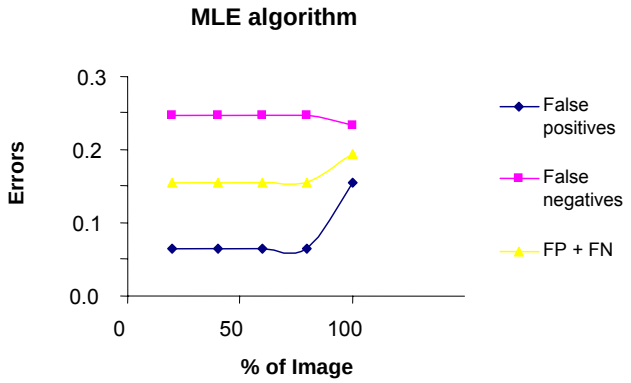


Figure 5. Percentage of errors when considering only the third RGB component for each pixel, using the MLE algorithm and the Euclidean Weighted distance

7. CONCLUDING REMARKS AND WORK IN PROGRESS

We have presented experimental results concerning the efficient detection of child pornography in the network infrastructure. We tested two estimator algorithms using four well-known linear classification mechanisms. The preliminary experiments show that our proposed systems may be able to detect the transmission of child pornography files through routers, with allowable errors.

Future work includes the actual setup of a peer-to-peer network in our internal laboratory with traffic diversity, and the software implementation of the algorithms in an intermediary IP router, which will help obtain more accurate results as to the efficiency of the detection mechanisms. Furthermore, it would be interesting to process random portions of images, as opposed to the upper portion, as we did. In addition, we are interested in evaluating the performance of our system when tested with non-child pornography datasets consisting of legal pornography (which may be harder to distinguish from child-pornography images). Another possible avenue for future investigation is to incorporate other classification and dimensionality reduction techniques. Classifiers such as support vector machines and Chernoff-distance based dimensionality reduction methods would improve classification accuracy, while performing fast classification at the expense of slowing the training phase.

ACKNOWLEDGEMENTS

This research is being supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Chilean National Council for Technological and Scientific Research (CONICYT). We acknowledge Munish Chopra for his contribution in the earliest stages of this paper and valuable comments on subsequent versions. We would like to thank the Toronto Police for facilitating the data sets for our experiments.

REFERENCES

[1] Aiken, J., 2005. No one fix for Net porn and kids. *CNN.com*. Available online: http://cnnstudentnews.cnn.com/2002/TECH/internet/05/02/youth.internet.porn/ [Accessed: June 24, 2005].

[2] Allen, W.H., 2005. Computer Forensics. *IEEE Security & Privacy*. Vol. 3, No. 4, pp. 59-62.

[3] Alorie, G. et al. Cnet News.com: Do web filters protect your child? January 24, 2006. Available online: http://news. com [Accessed: July 12, 2006].

[4] Apple Computer Inc., 2005. Family. *Apple.com*. Available online: http://www.apple.com/macosx/features/family/ [Accessed: June 24, 2005].

[5] Bajai Inc., http://www.bajai.com/home.html [Accessed: June 28, 2005]

[6] Bill C-2, 2006. Amendments to protect children and other vulnerable persons. *National Child Exploitation Coordination Centre (NCECC)*. Available online: http://ncecc.ca/backgrounder_c2_e.htm .[Accessed: May 16,2006].

[7] Butt, D., 2005. Working together for child safety: Collaboration and new technologies can help fight online sexual abuse of children. Available online: http://www. microsoft.com/issues/essays/2005/04-06childsafety.asp [Accessed: May 20, 2006].

[8] Chopra, M. et al., 2006. A source address reputation system to combating child pornography at the network level. In *Proceedings of the IADIS International Conference on Applied Computing (Outstanding Paper Award)*. San Sebastian, Spain, pp. 472-477.

[9] Duda, R. et al., 2000. *Pattern Classification (2nd Edition)*, Wiley-Interscience.

[10] Internet Online Summit: Focus on children. December 1-3 1997. Available online: http://www.enough.org/summit/ whitepaper.htm#Anchor.14 [Accessed: May 23, 2006].

[11] http://www.kazaa.com

[12] Keila, P.S. and Skillicorn, D.B., 2005. Detecting unusual and deceptive communication in email. Technical Report, School of Computing, Queen's University.

[13] Kherfi, M.L. et al., 2004. Image retrieval from the World Wide Web: Issues, techniques, and systems. *ACM Computing Surveys (CSUR)*, Vol. 36, No. 1, pp. 35-67.

[14] http://www.limewire.com

[15] Leclair, N., 2006. Brief overview of key cases in Canada. *Canadian Department of Justice*. Available online: http://ncecc.ca/case_law_e.htm [Accessed: May 17, 2006].

[16] Legislative Summaries, 2006. *Library of Parliament- Parliamentary Information and Research Services*. Available online: http://www.parl.gc.ca/common/Bills_ls .asp?Parl =38&Ses=1&ls=C2 [Accessed: May 17, 2006].

[17] Lyu, S. and Farid, H., 2006. Steganalysis using higher-order image statistics. *IEEE Transactions on Information Forensics and Security*, Vol. 1, No. 1, pp. 111-119.

[18] Matrawy, A. et al., 2005. Mitigating network denial of service through diversity-based traffic management. In *Proc. 3rd Intl. Conf. on Applied Cryptography and Network Security (ACNS)*, New York, USA, June 7–10.

[19] Moore, A.W. and Papagiannaki, K., 2005. Toward the accurate Identification of network applications," *Proc. Passive & Active Networks Measurement Workshop*, Boston, USA, 2005.

[20] Ohzahata, S. et al., 2005. A traffic identification method and evaluations for a pure P2P application. In *Proc. Passive & Active Networks Measurement Workshop*, Boston, USA,

[21] Oommen, B.J. and Rueda, L., 2005. Stochastic learning-based weak estimation of multinomial random variables and its applications to pattern recognition in non-stationary environments. *Pattern Recognition*, Vol. 39, 2006, pp. 328-341.

[22] Oommen, B.J. and Rueda, L., 2005. On utilizing stochastic learning weak estimators for training and classification of patterns with non-stationary distributions. In *Proc. of the 28th German Conference on Artificial Intelligence*, Koblenz, Germany, Springer, LNAI 3698, pp. 107-120.

[23] Paxson, V., 1998. Bro: A system for detecting network intruders in real-time. In *Proc. Of the 7th Annual USENIX Security Symposium*.

[24] Reviewing the Reviews: Parental Control Software. March 2006. Available online: http://www.consumersearch.com/ www/software/parental-control-software/fullstory.html [Accessed: June 15, 2006].

[25] Schell, B.H. et al., 2006. Cyber child pornography: A review paper of the social and legal issues and remedies—and a proposed technological solution. *Aggression and Violent Behavior, A Review Journal* . Elsevier. In Press.

[26] Shanmugasundaram, K. et al., 2004. Payload attribution via hierarchical Bloom filters. In *Proc. Of the ACM CGS*.

[27] Shepherd, M., 2000. Content filtering technologies and Internet service providers: Enabling user choice. Available online: http://users.cs.dal.ca/~shepherd/filtering/ISPweb. htm [Accessed: May 26, 2006].

[28] Tand, C. et al., 2003. Peer-to-peer information retrieval using self organizing semantic overlay networks. In *Proc.of SIGCOMM*.

[29] Theodoridis, S. and Koutroumbas, K., 2006. *Pattern Recognition (3rd Edition)*. Elsevier Academic Press.

[30] Wang, K. and Stolfo, S.J., 2004. Anomalous payload-based network intrusion detection. In *Proc. of the 7th Int'l Symp. on Recent Advances in Intrusion Detection (RAID 2004)*, Sophia Antipolis, France, September 15–17.

[31] Watt, N., 2006. Dutch court lets paedophile party contest country's general election. Availble online: http://www.guardian.co.uk/international/story/0,,1822972,00.html#article_continue [Accessed: June 23, 2006].

[32] Wikipedia, http://en.wikipedia.org/wiki/Winny [Accessed: June 24, 2005]

[33] Zuev, D. and Moore, A., 2005. Traffic classification using a statistical approach. In *Proc. Passive & Active Networks Measurement Workshop*, Boston, USA.

APPENDIX: PARAMETERS USED IN THE WEIGHTED EUCLIDEAN METRIC

In this section we justify the parameters used in the Weighted Euclidean metric and explain why this metric's results are more accurate in most cases (see Sec. 6).

Let V_p be the feature vector for child pornography images $V_p = (v_{p0}\ v_{p1}, v_{p2}, \ldots, v_{pi}, \ldots, v_{pn})$ and $\boldsymbol{\sigma}^2 = (\sigma_0^2, \sigma_1^2, \ldots \sigma_i^2 \ldots, \sigma_n^2)$ be the standard deviation vector, where σ_i^2 is the standard deviation of v_{pi}.

The standard deviation σ_i^2 for the feature vector v_{pi} that is important for a child pornography image will be smaller than the standard deviation of the feature that is non-important for a child pornography image. Ideally, a feature that is not important for the child pornography images will be important for non-child pornography images.

Thus, in the feature vector $v_p = (v_{p0}\ v_{p1}, v_{p2}, \ldots, v_{pi}, \ldots, v_{pn})$, some of the features are important for the child pornography images and the rest of them for the non-child pornography images (upon certain threshold criteria). Now, let $v_{ps1}\ v_{ps2}, v_{ps3}, \ldots. v_{psk}$ denote the features that are important for child pornography images and $v_{pt1}, v_{pt2}, \ldots. v_{ptm}$, the features that are

important for non-child pornography images (where $k + m = n$), and v_{psx} is different from v_{pty}, for all $r = 1, \ldots, k$, and all $s = 1, \ldots, m$).

Let us assume that $k < m$ which means that the number of features important for the child pornography images is less than the number of features important for non-child pornography images. Thus, let us rewrite the Weighted Euclidean Distance equation:

$$d_{WE}(V', V_p) = \sqrt{\sum_{r=1}^{k} w_{sr}(v'_{sr} - v_{psr})^2 + \sum_{r=1}^{m} w_{tr}(v'_{tr} - v_{ptr})^2} \qquad (1)$$

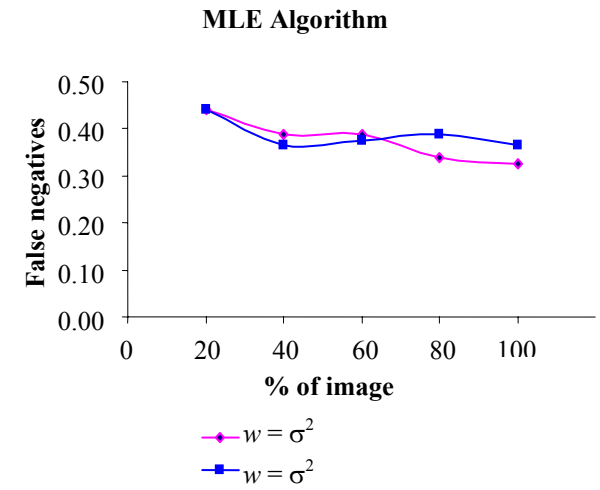
If we use as weighted factor $w_{ss} = 1/\sigma_t^2$, then we are expecting to have a bigger weighted factor for features that are important for child pornography than for non-pornography images. Conversely, when we use $w_s = \sigma_t^2$, actually we have used a bigger weighted factor for features that are important for non-child pornography images. The three figures below show the results for these two weighted factors.

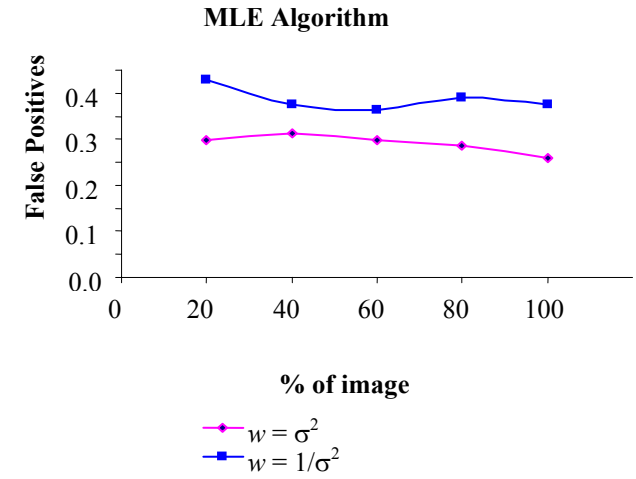
Since the false positive rate is lower for the weight factor $w_s = \sigma_t^2$ than for $w_s = 1/\sigma_t^2$, we are achieving a smaller number of errors in detecting non-child pornography images when using $w_s = \sigma_t^2$. In this case, the values of w_{tr} in the second summation of Eq. (1) are greater than w_{sr} in the first summation, and since $k < m$, we can say that the second summation dominates in calculating the distance $d_{WE}(V', V_p)$.

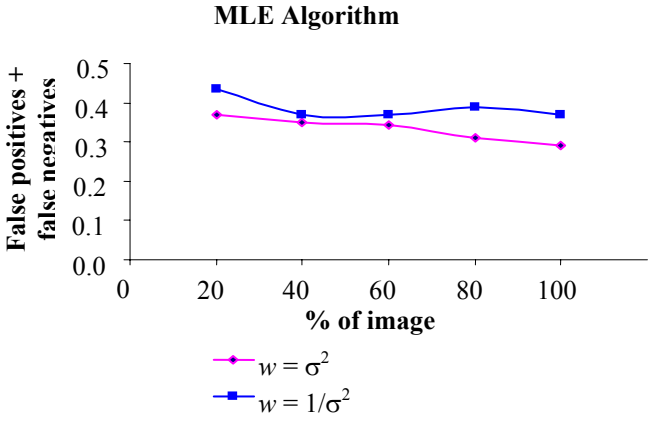
The false negative rate is almost the same for both weighted factors. This rate looks a little bit lower for $w_s = 1/\sigma_t^2$ than for $w_s = \sigma_t^2$ when the processed image percentage is lower. The system commits almost the same number of errors in detecting child pornography images. In this case the w_{sr} in the first summation of Eq. (1) is greater than w_{tr} in the second summation but since $k < m$ we can say that both summations have almost the same weight in determining the distance $d_{WE}(V', V_p)$.

Overall, the error rate is lower when using $w_s = \sigma_t^2$ than when using $w_s = 1/\sigma_t^2$. Therefore, the system is more accurate in classifying an image as child pornography or non-child pornography when using the weighted factor $w_s = \sigma_t^2$ than when using $w_s = 1/\sigma_t^2$. Ironically, to detect child pornography images (or image portions), it is best to try to detect non-child pornography images (or portions) and consequently, those images (or portions) that are not detected will be deemed as child pornography.

In summary, assuming the classes are normally distributed and the features are independent, then the classification rule as per Eq. (1) will result in an optimal (in the Bayesian context) classifier. That is, it is the classifier that assigns the packet to either child pornography or non-child pornography depending on the Mahalanobis distance from the class mean. Then, the larger the value of w_s, the more the features in the first summation of Eq. (1) will contribute to classify the sample. This rule, a particular case of the general scenario for normally distributed classes ([9, pp. 41-45]), minimizes the probability of classification error.


MLE Algorithm
False negatives
0.50
0.40
0.30
0.20
0.10
0.00
0
20
40
60
80
100
% of image
w = σ2
w = σ2


MLE Algorithm
False Positives
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
% of image
w = σ2
w = 1/σ2


MLE Algorithm
False positives + false negatives
0.5
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
% of image
w = σ2
w = 1/σ2